

Analysing spatiotemporal sequences in Bluetooth tracking data

Matthias Delafontaine^{*1}, Mathias Versichele¹, Tijs Neutens¹, Nico Van de Weghe¹

¹ CartoGIS Cluster, Department of Geography, Ghent University
Krijgslaan 281 WE12
9000 Ghent (Belgium)

^{*} Corresponding author: Matthias.Delafontaine@UGent.be

Abstract

The use of Bluetooth technology as a technique to collect data about the movement of individuals is increasingly gaining attention. This paper explores the potential of sequence alignment methods to analyse data obtained from Bluetooth tracking. To this end, an empirical case study is elaborated which applies sequence alignment methods to examine the behavioural patterns of visitors tracked by Bluetooth at a huge trade fair in Belgium. The results and findings underline both the validity of Bluetooth tracking to collect data from visitors at mass events, as well as the ability of sequence alignment methods to extract insightful information from sequences within such data.

1. Introduction

This paper will use sequence alignment methods to analyse patterns within tracking data obtained from Bluetooth sensing. Although existing as a communication technology since the mid-nineties, Bluetooth has only recently been employed for the tracking of individual movement [1-7]. Despite its limited positional accuracy, Bluetooth tracking is a low-cost alternative for true location-aware technologies. A major advantage of this technique is that it allows for the distinction of tracked subjects at the individual level. This is because Bluetooth-enabled devices broadcast a unique MAC (48-bit physical address). Furthermore, due to its widespread standard integration in nowadays personal wearable devices such as cellphones, PDA's and headsets, Bluetooth allows for unannounced tracking, i.e. tracking of subjects that are not aware of being tracked. It therefore offers researchers the valuable potential to conduct unbiased experiments and gather uninfluenced observations of a mass of individuals.

In this paper, we consider the most common approach to employ Bluetooth technology as a tracking system. It consists of a number of Bluetooth access points, henceforth denoted as nodes, installed at fixed strategic locations throughout the area of interest. Each node continuously searches for nearby devices. Whenever a Bluetooth-enabled device enters the radio range of a node, its MAC address is logged, such that the presence of devices at nodes can be recorded along the time line. From these records, the trajectory of an individual may be approximated as the spatiotemporal sequence of node observations of the device (s)he is carrying. In addition to this basic tracking system, optional supplementary attributes may be logged such as the device class¹ and its user-friendly name², although these might demand additional lookup time. To date, most Bluetooth tracking projects documented in the literature have relied on this concept (e.g. [3-6]). On the other hand, apart from being robust and plain, the concept is attractive due to its easy and low-cost implementation which requires merely a number of Bluetooth dongles³, computational units and storage units. Furthermore, the approach is efficient in its passive data collection as it does not set up true connections with devices, and thereby avoids any interaction with the individuals being tracked.

In the large body of research on movement behaviour, considerable work has been dedicated to the definition and extraction of patterns from movement data (e.g. [8-11]). Most of these approaches stem from a cross-pollination of GIScience, computational geometry, knowledge discovery in data bases, data mining, spatial cognition, and artificial intelligence [8, 12, 13]. However, much of these techniques may not be suitable to analyse Bluetooth tracking sequences. This is because Bluetooth tracking sequences may be incomplete or inconsistent due to data failure of the nodes (e.g. signal obstructions, data loss) and tracked devices (e.g. limited battery lives, disabled by the carrier) on the one hand, and due to the limited coverage of the study area in terms of node radio ranges on the other hand. REMO [14, 15], a generic geographic knowledge discovery approach to describe relative motion patterns through a matrix, for example, would not be a suitable formalism to represent and explore Bluetooth tracking sequences as it would require the location of each device to be known at regular time stamps. Another example is the Qualitative Trajectory Calculus [16, 17], which, despite its potential to handle incomplete information, is not eligible for handling Bluetooth sequences as it builds on higher level motion attributes such as motion azimuth and velocity.

This paper will explore the potential of sequence alignment methods (SAM) for the extraction of patterns within Bluetooth tracking sequences. SAM is a relatively new technique in the research field focusing on movement patterns. In the next section, we will briefly highlight the basics of SAM as well as related background work. Then, in section 3, we will apply SAM to analyse Bluetooth tracking sequences gathered at a 5-day trade fair in Ghent (Belgium). Finally, conclusions are drawn in section 4.

2. Sequence Alignment Methods

¹ The device class is a 3-byte value that describes a device by a hierarchical classification, e.g. Phone: Cellular, Computer: Laptop.

² A user-friendly name is an arbitrary word or phrase most often configurable by the user.

³ A Bluetooth receiver integrated into a USB stick.

2.1. Background

Having a tradition in bioinformatics to measure the distance between DNA strings or protein strands [18], SAM was first applied in social science by Abbott [19] to analyse career patterns. In turn, Abbott's pioneer contribution has triggered an important body of SAM studies within sociology (see [20] for an overview). From then on, SAM has been considered a promising methodology to analyse the sequential aspects of human space-time activities, which is reflected, among others, through contributions by Wilson, [21-23], Joh et al. [24-27], and Shoval et al. [28, 29].

Within the abundant research on human activity and travelling behaviour, SAM is usually applied to data collected by means of questionnaires, activity-travel diaries and position-aware devices. The application of SAM to data obtained from passive wireless tracking systems has, until present, not been scrutinised. An exception is the recent work of Choujaa and Dulay [30, 31] who consider activity sequences inferred from cellphone data. However, they employ SAM as a novel approach to predict gaps in the activity logs, rather than to analyse these logs.

2.2. Methodology

Sequence alignment is the process of equating two or more sequences of elements of a well-defined universe using a set of eligible operations [18]. Sequence alignment methods (SAM) seek for optimal alignments by employing dynamic programming algorithms to either maximise a similarity measure, or to minimise a distance measure [23]. This distance measure is usually referred to as Levenshtein distance [32, 33] or biological distance [28, 34]. There exist two categories of SAM algorithms. *Global alignment* methods force the alignment to span the entire length of the sequences, while local alignment methods focus on the similar parts within sequences that may differ significantly overall [31].

The conventional operations eligible for a pairwise alignment, i.e. the alignment of two sequences, are *identity*, *substitution*, *insertion*, and *deletion*. As they always occur together, the latter two operations are known as *indels* and are accommodated by gaps in one of both sequences. Sequences are usually represented as a string of elements consisting of one or more characters. A pairwise alignment of two single-character strings 'Bluetooth' and 'Blåtand'⁴ is illustrated in Figure 1. A multiple alignment, i.e. an alignment of three or more sequences, is usually approximated by a procedure of multiple pairwise alignments, known as progressive alignment [35].

B	l	u	e	t	o	o	t	h	
B	l	å	–	t	a	n	d	–	
✓	✓	✗	–	✓	✗	✗	✗	–	

✓ identity
✗ substitution
– indel

Figure 1 – Pairwise alignment

⁴ Bluetooth is named after the Danish king Harald Blåtand (940 – 981 A.D.).

To determine whether an alignment is optimal, the operations have to be weighted by a priori defined similarity scores. Typically, some additive scoring scheme is adopted in which the *identity* operation represents the highest similarity and is thus given the highest score. *Substitutions* are mostly associated to zero scores and *indels* to penalties (negative scores). However, depending on the nature of sequenced elements, combination-specific substitution scores (or indel penalties) may be useful. For instance with respect to the alphabet characters in the example, from an etymological-linguistic point of view, the *t-d* substitution might be assigned a higher similarity score than the *o-n* substitution. Specific similarity values are usually described by a scoring matrix which contains all pairwise substitution scores.

Contrary to traditional measures such as Euclidean, Manhattan, or Hamming distances, Levenshtein distances systematically capture the entire sequential dimension to assess the similarity among two sequences [28]. This is the principal advantage of SAM with respect to other methods. In addition, the alignment process allows for discovering hidden patterns buried within the dataset [21]. This is a particularly valuable characteristic within the context of this paper, given the frequent gaps in Bluetooth tracking logs.

According to Shoval and Isaacson [28], two types of analysis can be conducted on the basis of SAM. The most common one is an analysis of clusters of similar sequences and/or representative sequences. Another possibility consists of detecting hypothetical behavioural patterns within the sequence data at hand. The former use of SAM will be considered in the next section of this paper.

3. Case study

In this case study we will apply sequence alignment methods to analyse the behavioural patterns of visitors tracked by means of Bluetooth at the Horeca Expo in Ghent (Belgium). The Horeca Expo is the most important annual trade fair for the hotel and catering industry in Belgium, and it is particularly well-chosen as a setting for the examination of visitor movement patterns for several reasons. In the first place, the fair is a well-organised and controlled indoor event which is exclusively accessible for paying visitors, exhibitors and crew members. This strongly limits the potential interference and data noise due to all kinds of passers-by devices out of the study scope, which is, for instance, less evident in outdoor environments (e.g. [3, 7]). Secondly, we were allowed to passively track participants without their prior knowledge, such that the experiment is by no means biased in that sense. In addition, the daily variation and extent of additional smaller events that may cause temporary deviant behaviour of visitors during the fair is strongly limited. The data collection, preparation and results are discussed in depth in the remainder of this section.

3.1. Data collection

The data for this case study have been collected during the 21st edition of the Horeca Expo (November 22-26, 2009). This edition has counted 53 146 visitors, most of them being professionals in the catering industry, for 607 exhibition stands. The Horeca Expo takes place at the Flanders Expo exhibition centre in Ghent (Belgium). The centre has eight exhibition

halls over an area of about 56 000 m² (Figure 2). Each hall groups exhibition stands of a specific theme (e.g. hall 1: breweries, hall 5: kitchen contractors). 22 Bluetooth nodes, denoted A – T⁵, have been discreetly installed throughout the entire site. The nodes are equipped with power class 2 Bluetooth dongles which are developed to cover a radio range of about 20m, although experiments have shown that this range may vary substantially, among others due to indoor reflections. Given this presumption, it follows that on the one hand the study area is not completely covered by all nodes, and on the other hand some node pairs have an overlap in their covered areas (Figure 2).

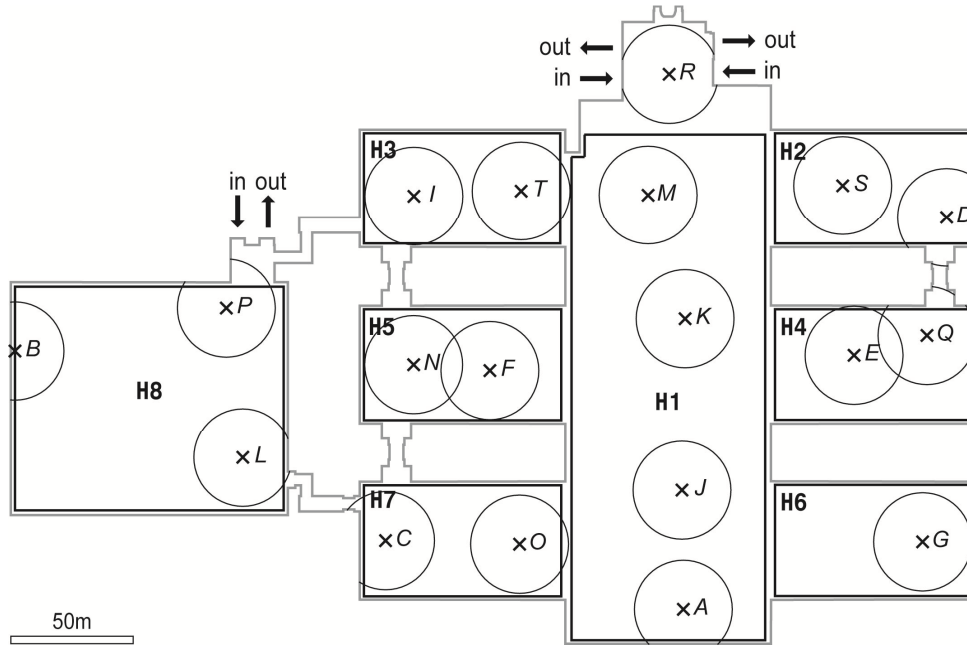


Figure 2 – Schematic map of Flanders Expo with indication of entrances and exits for visitors (arrows), exhibition halls (H1-H8, black rectangles), and Bluetooth nodes (A-T, x-marks) with 20m radio range (black circles)

The Bluetooth nodes continuously scan for nearby devices and log all discovered MAC addresses with the timestamp of discovery. Over the entire 5-day course of the fair, 14 498 unique devices have been observed, most of which are mobile phones and the like (95%) (Figure 3). Although, at most 2%⁶ of the observed devices are not wearable (e.g. desktop computers), these will be detained for further analysis since their tracking logs are not expected to reflect visitor movements. 89% of all devices have been observed only on one day (Figure 4), which suggests a large majority of one-day participants. In terms of unique devices per day, the dataset consists of 20 148 device-days. A histogram of device-day duration, i.e. the duration between the first and last node observation of a device on a day, is depicted in Figure 5. Two notable remarks can be drawn. On the one hand, almost 20% of the device-days have been observed for less than fifteen minutes. This can be explained among others by a quick disabling of devices of persons entering the fair and by short Bluetooth-enabled episodes of devices of people who intentionally make use of the Bluetooth functionality. Since this case study aims to analyse visitor behavioural patterns,

⁵ Node H has been left out as it is located out of the study area in this case study.

⁶ Including the devices for which the class is unknown.

such fragmented device-day observations can be considered unrepresentative and have therefore been excluded. On the other hand, over 10% of the device-days have observations that covering over eight hours, which is about the daily opening duration of the fair. Since these devices most probably accrue to exhibitors, crew members and/or are non-wearable, they have been excluded as well.

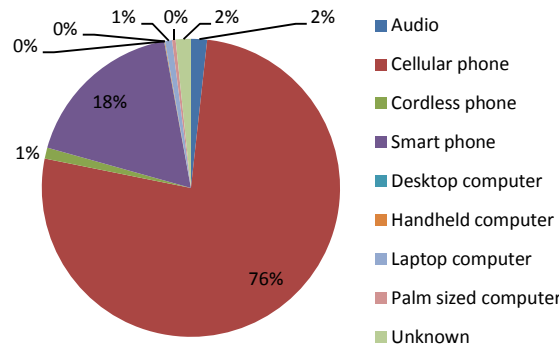


Figure 3 – Distribution of Bluetooth device classes across observed devices

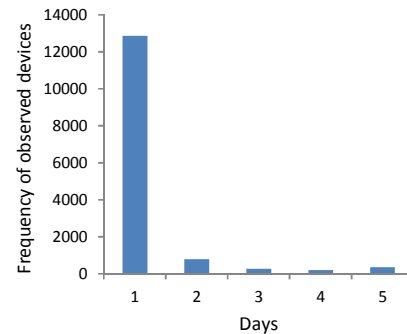


Figure 4 – Histogram of observed days per device

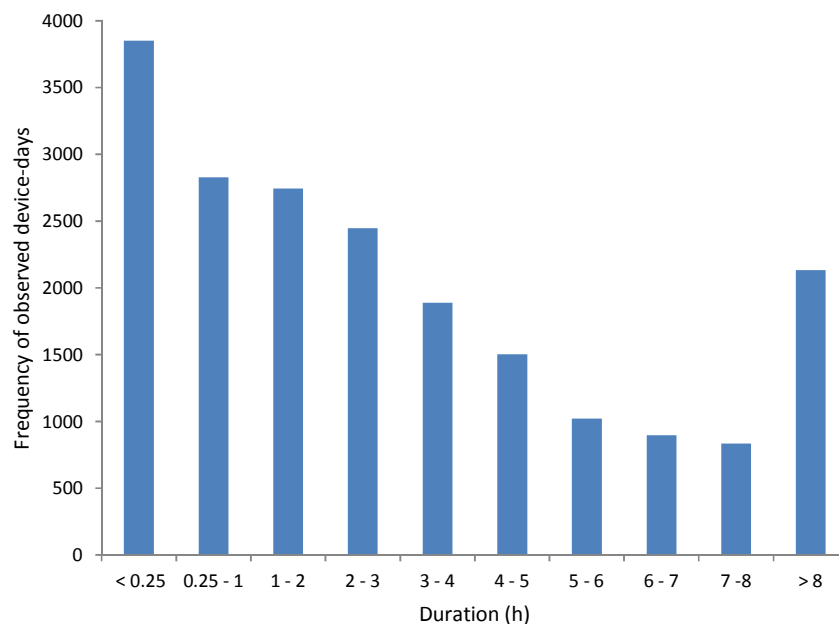


Figure 5 - Histogram of device-day duration

3.2. Data preparation

For each remaining device-day we have determined the chronological sequence of node observations. To filter for noise in the data, subsequent observations by the same node that are less than one minute apart have been concatenated to one observation lasting over the entire interval. Some additional preparative steps have been taken to extract representative sequences for visitors and to exclude as much as possible the sequences of exhibitors, crew members and outlier sequences. The following restrictions have been imposed:

- The first and last observations in the sequence are observed at node *P* or *R* which are located near the visitor entrances and exits (Figure 2);
- The time span of a sequence is within the official opening hour intervals of the fair, i.e. each day from 10:30 a.m. to 7:00 p.m.;
- The time gaps in between two subsequent observations in the sequence have a maximum duration of 15 minutes;
- The sequence contains observations of at least eight different nodes.

Further, the observation sequences that respect the above restrictions have been transcoded to single-character sequences to facilitate sequence alignment. To this end, a temporal unit of 3 minutes has been postulated as being the minimum duration for visiting a certain location within the fair. Hence, the observation sequences have been divided into 3-minute episodes, each of which has been allocated a character according to the following rules:

- If more than 50% of an episode is covered by observations of the same node, the node's character is allocated to the episode;
- If more than 50% of an episode is covered by observations of two nodes, the character of the node which observations cover the greater share is allocated to the episode;
- If an interval has observations of three or more nodes, a character *V* is allocated to the interval;
- In all other cases a gap character (-) is assigned.

Figure 6 presents some of the resulting sequences. The interpretation of sequence characters is as follows. A node character represents a visiting event in the neighbourhood of the corresponding node; a *V* character represents a travelling episode, i.e. a visitor travelling through the fair (e.g. in between two visiting events); and gaps represent the unknown information. Note that SAM are – more than any other methodology – able to handle gaps which are interpreted as indel operations (section 2.2). Given the above constraints and the strategic dispersion of nodes across the study area (Figure 2), it is probable that visitors remain near to the node of their last observation during gaps. As the interpretation of gaps and *V* episodes may depend on neighbouring characters, sequences consisting for more than 50% of gaps or *V* episodes have been excluded.

S	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	1
T	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	10	1
V	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3

Figure 7 – Sequence alignment scoring matrix

510 sequences were found to validate the restrictions imposed by the data preparation (section 3.2). Using the yet specified similarity scores and penalties, a multiple alignment of these sequences has been generated within the ClustalTX software package [23] by means of a progressive alignment procedure which consists of (i) a pairwise alignment of all sequence pairs (i.e. 129 795 pairs) using a local alignment algorithm (Smith-Waterman [36]), (ii) a neighbour-joining process (Saitou-Nei [37]), and (iii) a multiple alignment using a global alignment algorithm (Needleman-Wunsch [38]). The neighbour-joining process aims to structure the sequence data by joining similar sequences on the basis of their pairwise alignment score such that a guide tree is derived which determines the optimal order for adding sequences to the multiple alignment by proceeding from the leaves to the root of the tree.

3.4. Results

Three results are obtained from the threefold sequence alignment process described in section 3.3: (i) a square matrix with pairwise alignment scores, (ii) a neighbour-joining guide tree, and (iii) a multiple alignment. The alignment matrix is the most raw and low-level result which will not be further considered as the information it contains is captured by the other results.

The guide tree (ii) obtained from the neighbour-joining process (section 3.3) is shown in Figure 8. It totals 509 hierarchical clusters of similar sequences. The clusters observed in this guide tree may assist in the determination of a typology of different visitor behavioural patterns. The number of members in a cluster can thus be considered an indicator for the importance of the corresponding behavioural pattern. Sequences in smaller clusters, however, tend to have more elements in common. In SAM literature regarding activity patterns, it is usually considered up to the analyst to determine the number and interpretation of clusters in the guide tree. This can be facilitated by means of the multiple alignment (iii). To enable a visual exploration of patterns in the multiple alignment, we have sorted the aligned sequences according to the leave order of the guide tree. In addition, the node characters in the multiple alignment have been colour coded according to the exhibition hall where they are located. A fragment of this representation is displayed in Figure 8. It illustrates a clear pattern of related sequences with predominant episodes at the exhibition halls 8, 7 and 1 (see further cluster 1.1).

Within the guide tree, three major clusters (Figure 8, 1-3) can be observed at the top of the hierarchy. At this level, the aligned sequences hardly share common characteristics, if at all. On the basis of visual supervision of the sorted and colour coded multiple alignment (Figure 9) an exhaustive subdivision has been made into 21 non-overlapping subclusters (Figure 8, 1.1-3.8). For each subcluster the number of members and the shared pattern has been summarised in Table 1. In addition, the subcluster median and average sequences have been

listed in Table 2. The median and average sequences are representative sequences of a cluster [23]. In analogy to the homonymous descriptive statistics, these sequences respectively minimise the sum Levenshtein distances and the sum of squared Levenshtein distances to all other members of the cluster.

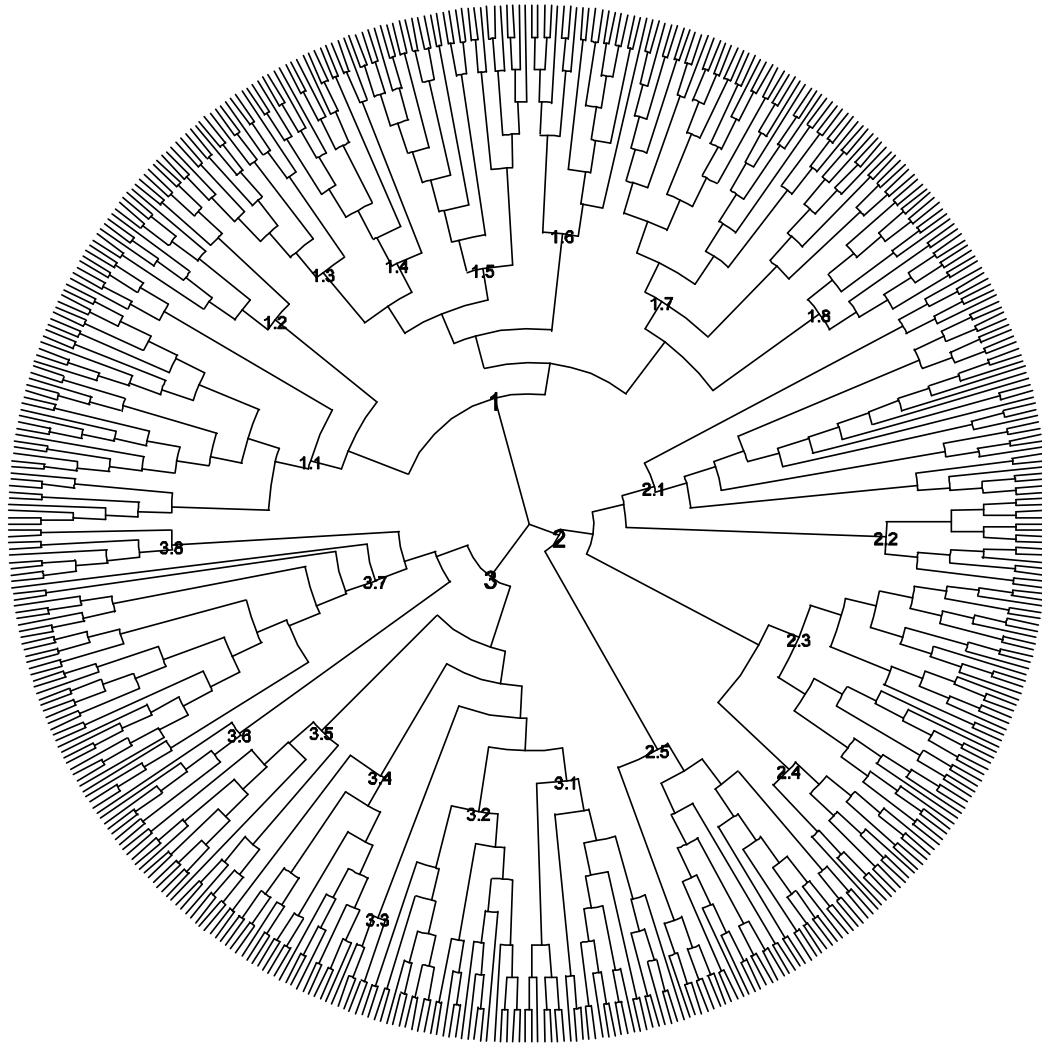


Figure 8 – Multiple alignment guide tree with clusters and subclusters labeled at their root node

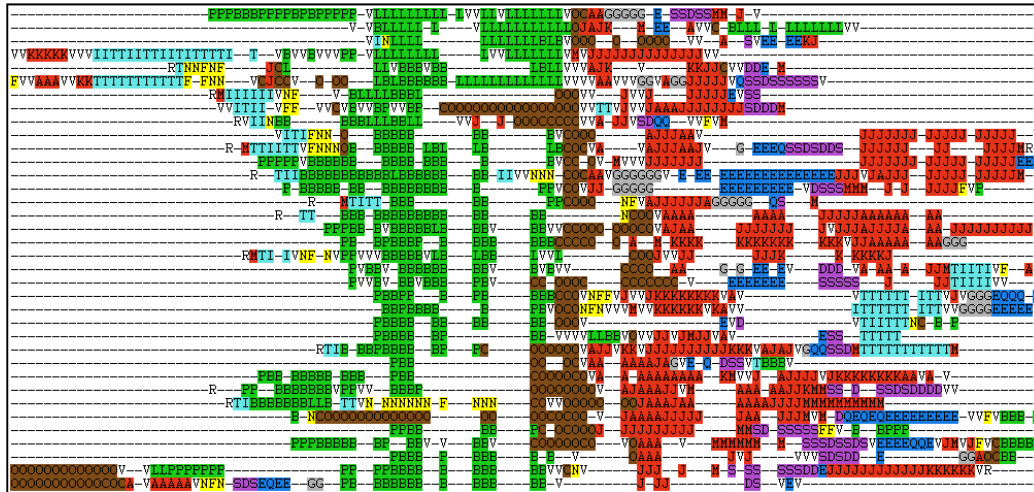


Figure 9 – Extract of the sorted and colour coded multiple alignment (colour legend: ■ hall 1, ■ hall 2, ■ hall 3, ■ hall 4, ■ hall 5, ■ hall 6, ■ hall 7, ■ hall 8)

Cluster	Members	Common pattern	Legend
1.1	47		<div> predominant episode frequent episode occasional episode disjunction of episodes </div> <div> 1 – 8 exhibition halls A – T Bluetooth nodes </div>
1.2	17		
1.3	17		
1.4	18		
1.5	29		
1.6	19		
1.7	49		
1.8	17		
2.1	38		
2.2	15		
2.3	44		
2.4	14		
2.5	34		
3.1	26		
3.2	26		
3.3	5		
3.4	22		
3.5	19		
3.6	7		
3.7	38		
3.8	9		

Table 1 – Number of members and common patterns per cluster. Pattern episodes are colour coded to hall location and annotated with hall numbers or node characters. Hollow episode symbols represent episodes at one of the eight exhibition halls.

The results in Table 1 and Table 2 reveal some interesting aspects about the behaviour of Horeca Expo visitors. First of all, they reflect a large heterogeneity of visiting patterns in terms of visit duration, the number of visited locations, and in particular the order of visiting

these. Notwithstanding that the fair can be entered and left from only two locations, the variety of tracking sequences emphasizes the lack of one or a few dominant spatiotemporal behavioural patterns of Horeca Expo visitors. Inferences can be made concerning the attractiveness of locations, although these might be misleading since not all exhibition halls have been equally covered by Bluetooth nodes (e.g. hall 6). The abundant hall 1 episodes reflect that the main hall is also the most important one in terms of visits, as could be expected given its size and central location. More than that, it can be observed that most sequences visit the main hall more than once, whereas other halls are usually visited once at most. 12 of the 21 common cluster patterns in Table 1 feature two disjoint episodes at hall 1, whereas none of them features repetitive visits of other halls (except for hall 7 in cluster 2.1). Thus, people most often tend to benefit maximally from their visit by passing as much locations as possible, thereby avoiding revisiting halls, which is inevitable for the main hall. Merely one cluster (2.2) seems to represent an exhaustive visit to the fair, i.e. calling at all exhibition halls. However, given that only shared episodes have been listed in Table 1, many other clusters may encompass such visits as well (e.g. see Table 2, Figure 9).

Median and average sequence

```
1.1 VTTVDDDDDDDDDDDDCCCA-VAAAAVNFN-SDSQQE--GG--PBBBDBB-DBBBB-V---I-IA---DSvTv
PBBB-DBBB-CCDDIJJJJJJJJ-MSDSSSSSFV-F--PPPP
1.2 --VTT-VTV--VSSDQEE-GG--A-IJJJJJJJJIPBB--VVV
RRRRRRRRR-V-EEEEEQEEEEEEEEEE-JAAAAAJJJJJJAAAAAAAJAA-V
1.3 VV-F-VTIS-VQCC--DCC-CCC--EEEEEEEEEQVDSV--VMAA-AA--A-AAA-AAA-RAAK
R--V-F-0000TTISSE-I-BPVV---M-
1.4 R-S-SEI--INN--JCCCCCCCCCCCCCCCCCCCC--VTTTVV-SPITV
VVVAAAAAIAAAAAAAAAAAV-TITIIITITVCCCPBBB-PBBB-BB-VVVVGGGG--FDDDDSDS---SSS-SDSvS
1.5 VTIIITIT-IT-FFDDV-VVVVVVVVGGAAAAAIAAAV-VFFV
MJJVNNTV-VGVAAAAA-VVDVDDCCV
1.6 R-MMMN--VV-V-VAAAAAJAAAAJCC--LILBLLIITITN-IJJJ-I-JJJJ-VMMMM
RRR-IA-AA-VV-VVVF-EQEEEDSSM-I-ITNVA-EQDDSS-SSSS-JJJVGE--000000VVIV--BBBEPQIVRRRRRR
1.7 RRTIITITI--IIT-IVNNNFVAJKKKKHVQEDDSS-EGGGGG-IV-BBB-BBLLV
RVTV-I-G--VV-VVVEQVDSVMIIIN-NNVCCVJJJJJV
1.8 RSSDDSVIVJJJJJJJJVVVVVVVVVVVMMMM00MMKKKKEE---EQEDV-R
PBBBPBBB-CCCVVVVVGGGVFFNVITVBBBBB---BBBBBBB
2.1 RVAAVVIV-M--M-AV-MITITITITIIIVPBBB-BBBB-BBBB-VV-000000000-EQV-SDSDS-VQIV-
VSS-DEVGGA-MMMJNVLEB-EVNVI
2.2 RVDSSS-DEEEEGVGGGVMMJAA-AV-V-BBLLIILIIIVBBB-VTIIITIN-V-FVIVJJJJJJ-JJ
RR-VS-DBSSS-SDDDVV-VGGAVVDDV0000CV---IITVDSDDSDDDSDSDS-
2.3 RSDSDD-D-DDDDDDDDDDDEEQ---EEEEEEEEEEEFJCCCFNIIIT-V
R--RMITT-FFNNVAAA-VTISSEEDDDDDDEEEEDQ0000Q00EEF--QE-AVNVVBBB-B-BBBB--V-V-GGGGGVGCAGVNVVVR-
2.4 RVGGGGGGGGG-DEFFNNV-V--ED--
```

2.5	
3.1	
3.2	
3.3	
3.4	
3.5	
3.6	
3.7	
3.8	

Table 2 - Median (top) and average (bottom) sequence per cluster (colour legend: ■ hall 1, ■ hall 2, ■ hall 3, ■ hall 4, ■ hall 5, ■ hall 6, ■ hall 7, ■ hall 8)

4. Conclusions

In this paper, we have explored the potential of sequence alignment methods to analyse data obtained from tracking individuals by means of Bluetooth. After a brief introduction on Bluetooth tracking and sequence alignment methods (SAM), an experimental case study has been presented on the sequence alignment analysis of spatiotemporal patterns of visitors tracked by Bluetooth nodes at the Horeca Expo fair in Belgium. The contribution of this work is original since, until present, SAM has not been applied to analyse Bluetooth tracking data. The importance of this study also relates to the growing attention paid to Bluetooth as a novel technology to track people at mass events. We have shown that, provided that the necessary steps have been taken to filter raw Bluetooth tracking data, SAM can be successfully adopted to analyse Bluetooth tracking sequences. The results of the case study have revealed some important and plausible insights about the behaviour of visitors at the Horeca Expo. In particular, the study has disclosed the existence of a large variety of visiting patterns especially with respect to the number of and order of visited locations. Notwithstanding this considerable heterogeneity, we have demonstrated the ability of SAM to detect and extract the sequential structures hidden in the tracking data. The vast majority of tracking sequences respects a reasonable chronological concatenation of visited locations, which in turn confirms the ability of, in essence, simple Bluetooth tracking systems to capture the spatiotemporal behaviour of large crowds of individuals at a mass event. The results of our study may be insightful to the planners and organisers of such events in keeping track of and exploring the behaviour of participants over the course of an event.

Despite the above contributions, some aspects still limit the potential of sequence alignment methods for the analysis of tracking data. Unlike the structure of nucleotides in a strand of DNA, spatiotemporal sequences within tracking data might differ very much amongst tracked individuals, both with respect to sequence composition as with respect to the number of elements (duration). In sequence alignment, the latter aspect may cause a large number of gaps, for which there is yet no consensus on their interpretation [35]. Shoval and Isaacson [28] recognize the lack of a solid method to assess the reliability of alignments, as well as the lack of knowledge on the impact of the spatial and temporal scale on the results. Other issues relate to the shortcoming of SAM as an exact science, or to quote Morrison [18]: “The basic problem with sequence alignment is that it seems to be more an art than a science”. For example, there is no consensus method or standard calibration procedure for the setting of sequence alignment parameters such as indel penalties. Regarding tracking data, even common practices are lacking in this respect. Future progress on these issues will enable more refined analysis configurations and support stronger and more detailed interpretations of alignment results.

Acknowledgments

The authors express their great gratitude to Artexis for allowing and supporting Bluetooth tracking at the Horeca Expo. The Research Foundation – Flanders is gratefully acknowledged for funding the research of Matthias Delafontaine and Tijs Neutens; the agency for Innovation by Science and Technology for funding the research of Mathias Versichele.

References

1. O'Neill, E., Kostakos, V., Kindberg, T., Schiek, A., Penn, A., Fraser, D., Jones, T.: Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape. *UbiComp 2006: Ubiquitous Computing* (2006) 315-332
2. Hermersdorf, M., Nyholm, H., Perkiö, J., Tuulos, V., Salminen, J., Tirri, H.: Sensing in Rich Bluetooth Environments. In: ACM (ed.): *WSW'06 at SenSys'06*. ACM, Boulder, Colorado, USA (2006) 5
3. Van Londersele, B., Delafontaine, M., Van de Weghe, N.: Bluetooth Tracking - a spy in your pocket. *GIM International*. Geomares Publishing (2009) 23-25
4. Fallast, K., Scholz, A., Ekam, H.W.: Sichere Basis für Verkehrsplanung: Erfassung von Fahrgastströmen via Bluetooth. *Der Nahverkehr* (2008) 72-75
5. Wasson, J.S., Sturdevant, J.R., Bullock, D.M.: Real-Time Travel Time Estimates Using Media Access Control Address Matching. *ITE Journal* **June 2008** (2008) 20-23
6. Bullock, D.M., Haseman, R., Wasson, J.S., Spitler, R.: Anonymous Bluetooth Probes for Measuring Airport Security Screening Passage Time: The Indianapolis Pilot Deployment. *Transportation Research Board Annual Meeting 2010* (2010) Paper #10-1438
7. Furbach, U., Maron, M., Read, K.: Information Systems for Spatial Metro. In: van der Hoeven, F.D., Smit, M.G.J., van der Spek, S.C. (eds.): *Street-level desires. Discovering the city on foot*. Delft University of Technology, Department of Urbanism, Delft, The Netherlands (2008) 74-79
8. Laube, P., Wolle, T., Gudmundsson, J.: Movement patterns in spatio-temporal data. In: Shekhar, S., Xiong, H. (eds.): *Encyclopedia of GIS* (2007)
9. Gudmundsson, J., van Kreveld, M., Speckmann, B.: Efficient Detection of Patterns in 2D Trajectories of Moving Points. *Geoinformatica* **11** (2007) 195-215
10. Dodge, S., Weibel, R., Lautenschutz, A.: Towards a taxonomy of movement patterns. *Inf Visualization* **7** (2008) 240-252
11. Dodge, S., Weibel, R., Forootan, E.: Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems* **33** (2009) 419-434
12. Gottfried, B., Aghajan, H.: Behaviour Monitoring and Interpretation - BMI: Smart Environments, Volume 3 Ambient Intelligence and Smart Environments. IOS Press (2009)
13. Miller, H.J., Han, J.: *Geographic Data Mining and Knowledge Discovery*. Blackwell Publishing Ltd (2008)
14. Laube, P., van Kreveld, M., Imfeld, S.: Finding REMO - Detecting relative motion patterns in geospatial lifelines. In: Fisher, P. (ed.): *Developments in Spatial Data Handling*. Springer-Verlag Berlin, Berlin (2005) 201-215
15. Laube, P., Imfeld, S., Weibel, R.: Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science* **19** (2005) 639 - 668
16. Delafontaine, M., Chavoshi, H., Cohn, A.G., Van de Weghe, N.: A Qualitative Trajectory Calculus to reason about moving point objects. In: Hazarika, S.M. (ed.): *Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions*. IGI Global (2010) 350
17. Delafontaine, M., Cohn, A.G., Van de Weghe, N.: Implementing a qualitative calculus to analyse moving point objects. *Expert Systems with Applications* **In Press, Corrected Proof**
18. Morrison, D.A.: Sequence Alignment: Methods, Models, Concepts, and Strategies. *Systematic Biology* **59** (2010) 363-365
19. Abbott, A.: Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology* **21** (1995) 93-113

20. Abbott, A., Tsay, A.: Sequence Analysis and Optimal Matching Methods in Sociology. *Sociological Methods & Research* **29** (2000) 3-33
21. Wilson, W.C.: Activity pattern analysis by means of sequence-alignment methods. *Environment and Planning A* **30** (1998) 1017-1038
22. Wilson, C.: Activity Patterns of Canadian Women: Application of ClustalG Sequence Alignment Software. *Transportation Research Record: Journal of the Transportation Research Board* **1777** (2001) 55-67
23. Wilson, C.: Activity patterns in space and time: calculating representative Hagerstrand trajectories. *Transportation* **35** (2008) 485-499
24. Joh, C.-H., Arentze, T.A., Timmermans, H.J.P.: Multidimensional Sequence Alignment Methods for Activity-Travel Pattern Analysis: A Comparison of Dynamic Programming and Genetic Algorithms. *Geographical Analysis* **33** (2001) 247-270
25. Joh, C.H., Arentze, T.A., Timmermans, H.J.P.: A position-sensitive sequence-alignment method illustrated for space - time activity-diary data. *Environment and Planning A* **33** (2001) 313-338
26. Joh, C.-H., Arentze, T., Hofman, F., Timmermans, H.: Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological* **36** (2002) 385-403
27. Joh, C.-H., Arentze, T., Timmermans, H.: Identifying Skeletal Information of Activity Patterns by Multidimensional Sequence Alignment. *Transportation Research Record: Journal of the Transportation Research Board* **2021** (2007) 81-88
28. Shoval, N., Isaacson, M.: Sequence Alignment as a Method for Human Activity Analysis in Space and Time. *Annals of the Association of American Geographers* **97** (2007) 282-297
29. Shoval, N., Auslander, G.K., Freytag, T., Landau, R., Oswald, F., Seidl, U., Wahl, H.-W., Werner, S., Heinik, J.: The use of advanced tracking technologies for the analysis of mobility in Alzheimer's disease and related cognitive diseases. *BioMed Central Ltd.* (2008)
30. Choujaa, D., Dulay, N.: Aligning activity sequences for continuous tracking of cellphone users. *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on* (2009) 1-6
31. Choujaa, D., Dulay, N.: Activity Inference through Sequence Alignment. In: Choudhury, T., Quigley, A., Strang, T., Suginuma, K. (eds.): *Location and Context Awareness*, Vol. 5561. Springer Berlin / Heidelberg (2009) 19-36
32. Schlich, R.: Homogenous groups of travellers. *10th International Conference on Travel Behaviour Research, Lucerne* (2003) 29
33. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* (1966) 707 -710
34. Bargeman, B., Joh, C.-H., Timmermans, H.: Vacation behavior using a sequence alignment method. *Annals of Tourism Research* **29** (2002) 320-337
35. Wilson, C.: Reliability of sequence-alignment analysis of social processes: Monte Carlo tests of ClustalG software. *Environment and Planning A* **38** (2006) 187-204
36. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. *Journal of Molecular Biology* **147** (1981) 195-197
37. Saitou, N., Nei, M.: The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4** (1987) 406-425
38. Needleman, S.B., Wunsch, C.D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48** (1970) 443-453